Usage of measured reverberation tail in a binaural room impulse response synthesis

Markovic, Milos; Olesen, Søren Krarup; Madsen, Esben; Hoffmann, Pablo F.; Hammershøi, Dorte

Published in:
Acustica United with Acta Acustica

Publication date:
2011

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
Markovic, M., Olesen, S. K., Madsen, E., Hoffmann, P. F., & Hammershøi, D. (2011). Usage of measured reverberation tail in a binaural room impulse response synthesis. *Acustica United with Acta Acustica*, *97*(Supplement 1), S 122.

Usage of Measured Reverberation Tail in a Binaural Room Impulse Response Synthesis

Miloš Marković, Søren Krarup Olesen, Esben Madsen,
Pablo Hoffmann, Dorte Hammershøi
Section of Acoustics, Department of Electronic Systems, Aalborg University,
Aalborg Ø, Denmark.

Summary

The aim of the modern communication technologies is an immersive experience. One of the applications that should provide the feeling of being together and sharing the same environment during the communication process is BEAMING. The goal of this paper is to improve audible spatial impression utilizing correct acoustical properties of the specific environments.

Binaural room impulse response (BRIR) synthesis represents one of the main tasks in the binaural auralization. When the BRIRs are synthesized, high order reflections (reverberation tail) are usually modeled statistically because of the high density of reflections. That can lead to metallic and unnatural sound. Also, room-specific sound envelopment feeling is lost. This paper investigates the possibility of using measured reverberation tails instead of the modeled one in BRIRs synthesis. Three cases are observed. In the first one, BRIRs measurement in a real room isperformed. In the second one, synthesizedBRIRs are used. BRIRssynthesisisrealizedusing the image-source method for the earlyreflections and the artificialreverberationalgorithm for the reverberationtail. The third case combines modeledearlyreflectionsfrom the second case and measuredlatereverberation-from the first one. All three cases are evaluated and comparedobjectivelybased on the obtained room acousticparameters as well as subjectively by listening tests.

PACS no. 43.55.Ka, 43.66.Pn

1. Introduction

Immersive experience represents one of the primary goals of the modern communication technologies. Users' needs in communication go beyond the services that give the possibility of long distance real time conversation. Communication with the feeling of being together and sharing the same environment is required [1].

A project named BEAMING[1]is currently addressing the issue of improving immersive communication interfaces. BEAMING is a collaborate research project where the goal is to give people (visitors) a real sense of physically being in a remote location with other people (locals) without physically travelling. Simultaneous streams of data from the destination site to the visitor's perceptual apparatus, and from the actions and state of the visitor to the destination site, cohere together to form a unified virtual environment representing the physical space of the destination in real-time, a destination that now includes the beamed people [2].

Spatial sound plays an important role if a concept of real sense of physically being in a remote location wants to be obtained. Different techniques are used for spatial sound rendering [3]. Most of them are based on the similar principle: modeling of sound field and reproducing. Modeling relies on knowledge of sound propagation behavior in acoustical spaces while reproduction involves binaural cues of the human hearing.

Binaural technology enables realistic listening simulation of a variety of room environments. This auralization technique allows users to listen and evaluate the acoustics of the environment without being physically present in it. Spatial sound is achieved by convolution of monaural sound (simulated or recorded) with Binaural Room Impulse Response (BRIR).

For the purpose of the real time binaural synthesis with the moving sources and receivers, measuring of the BRIRs becomes impossible. Instead,

[1]Being in Augmented Multi-Modal Naturally Networked Gatherings, a four year FP7 EU collaborative project (#248620), started on Jan 1st 2010.

different modeling software is used for BRIRs modeling. This requires sufficient amount of time for the calculations. The needed time increases with theroom size and with the number of reflections. Also, during the various simplifications, some acoustical properties can be lost.
This paper investigates the possibility of using a measured reverberation tail instead of the modeled one in BRIR synthesis. It relies on the fact thatthere are certain similarities among the reverberation tails measured in the same room [4]. This indicates that finite (relatively small) number of measured reverberation tails can be used for more realistic binaural synthesis. In that way, premeasured tails can be used in combination with early reflections that can be modeled in real time.

2. Binaural Room Impulse Response

A BRIR consists of a pair of impulse responses representing an acoustical transfer functions of two transmission paths from a sound source to two receivers located in the ears (human or artificial head) [5]. It contains information about sound behavior in the environment (reflections in the room), as well as human binaural hearing features. It is of most importance, for good auralization, to obtain a BRIR for the specific source-to-receiver path. The most accurate approach for existing room is to measure BRIR for each position of interest. However, there are different practical limitations like large number of measurements in the room, time and equipment requirements, room accessibility etc. As an alternative to this, BRIRs can be simulated by acoustic room simulation software.
BRIRs can be divided into three parts: direct sound, early reflections and reverberation tail. Early arriving sound is more important for spatial impression than the late one. This is because the human hearing has strong suppressive mechanisms that mask input in the time domain (“precedence effect”) [5]. Another phenomenon called “binaural echo suppression” points that the effects of reflection, reverberation and background noise are less noticeable than when listening with one ear only [5]. Because of these phenomena many auralization systems model the direct sound and early reflection as accurate as possible while approximating the reverberation tail [6].

2.1. Reverberation tail

When a room is excited by an impulsive sound signal, the sound in the room decays as a function of time [7]. After numbers of reflection, the sound in the room may be considered diffuse: all directions of propagations are equally probable and the average energy density is the same across the room. After some time, individual characteristics of each reflection cannot be detected.
Different criteria for the reverberation tail starting point determination are proposed (transition time). The fixed value of 50 or 80 ms was suggested by classical architectural acoustics [e.g. 8]. Determination of transition time based on reflection order is common in BRIR synthesis. Approaches based on the mean free path suggest determination of a mean distance of a sound ray between two reflections in a room [7]. The mean free path of a room is determined as:

$$\bar{l} = 4\frac{V}{S}, \qquad (1)$$

where V is the room volume and S is the room surface area. Indirectly, transition time can be determined by checking the diffuseness of the sound field. For that purpose, determination of reflection density has been proposed as [7]:

$$\frac{dN}{dt} = \frac{4\pi c^3}{V} t^2 . \qquad (2)$$

According to the criteria mentioned above, transition time for a small and medium size rooms is placed between 20 and 80 ms.
Human hearing system is not equally sensitive to the changes in the reverberation tail as to the changes in the early arriving sound energy. However, the reverberation tail does contribute to some sound properties such as spatial impression and acoustical quality in enclosures. Especially, influence of the late arriving lateral reflections to the listener envelopment (LEV) is pointed out [8]. Thus, it is important to simulate the reverberation tail accurately enough at least to reflect correct acoustical properties according to the rendered acoustical scenes.
Most of the proposed techniques for the BRIR synthesis use statistical model for the late diffuse reverberant energy. Using a statistical model for the late reflections (reverberation tail) some specific acoustical properties of the room can be lost and the result of auralization might not be comparable with the original room listening experience.

3. Methods of investigations

3.1. Measured BRIR

The BRIR measurements were done in a standard listening room at Aalborg University. The base of the room has a rectangular shape with dimensions

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

of 7,8x4,18m^2. Three of four walls are with the slope on the top and the ceiling is a 2,78m high. Several BRIRs were measured at different positions in the room, with a different receiver orientation.

The BRIR measuring system consisted of a PC, a general purpose PC-based acoustical measuring device (01 dB *Symphonie*), a power amplifier, an omnidirectional dodecahedral loudspeaker, and an artificial head, ears and torso (*Valdemar, AAU*) with a pair of measuring microphones built in, Figure 1.

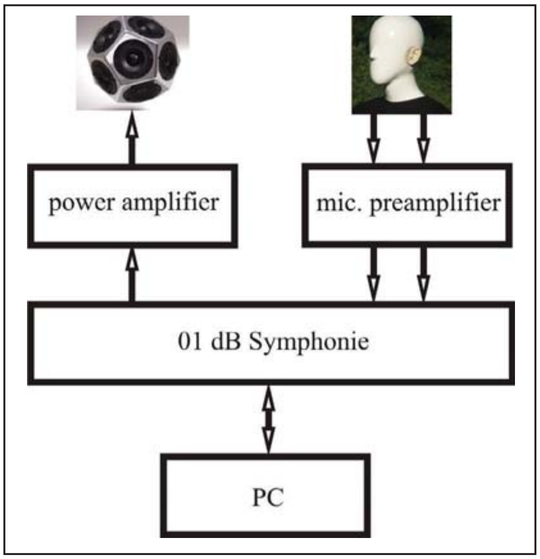

Figure 1. BRIR measuring system

The excitation signal was maximum length sequence of 16th order with 51,2 kHz sampling frequency. The computed BRIRs were derived from the average of 16 consecutive measurements by *Symphonie* system in order to improve noise immunity.

For the further analysis, BRIRs were measured in the middle of the room with source – receiver distance of 3,35m. Sound source was placed at 90^{O} to the right of the receiver and 0,35m above the line of the receiver's ears. The measured BRIR was resampled to the frequency of 44,1 kHz.

For the purpose of listening tests, measured BRIR is convolved with anechoic signals (speech and music). This gave a realistic auralization of the room where recording was performed. Convolution was performed using MatLab software.

3.2. Modeled BRIR

The modeled BRIR was obtained using the CATT-Acoustic software [9]. A 3D CAD model of the standard listening room was built on the basis of the geometric data, Figure 2. Data of the room surfaces (material properties) were assigned in the software *Prediction module*. Positions of the source and receiver, as well as receiver orientation were picked to follow their mutual relation from the measurements. Based on all provided information, full detailed echograms were obtained. To preserve high early part detail direct sound, first order diffuse and specular reflections, and second order specular reflections were handled deterministically by the image-source model (ISM). For the late reflections, *Prediction module*uses "randomized tail-corrected cone-tracing" (RTC) technique [9]. In a *Post-processing* module, obtained echograms were processed using the head related transfer function (HRTF) library which is included in the software. In that way, modeled BRIR is obtained. It was used for later room acoustic parameters comparison and evaluation.

Modeled BRIR was processed in the same way like measured BRIR. The same anechoic recordings were used here. The result was subjectively evaluated by listening test.

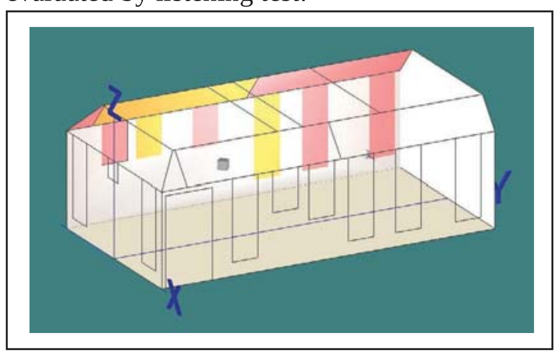

Figure 2. 3D CAD model of the standard listening room: surfaces with different material properties

3.3. Simulated BRIR

For the purpose of analysis, measured and modeled BRIRs are combined in the way that measured reverberation tail are added to the early reflections from the modeled BRIR. In that way, a simulated BRIR is obtained. The simulated BRIR was used for the room acoustic parameters comparison. After the convolution with the anechoic signals (the same signals have been used in measurements and modeling), subjective evaluation by listening test was performed.

Different concatenation points are chosen according to the proposed transition times mentioned before. Thus, concatenation time of 20, 40, 60 and 80 ms were investigated. These numbers indicate the length of modeled early reflection followed by measured reverberation tails. The concatenation process includes energy preservation and cross-fading at the junction of the concatenation.

3.3.1. Energy preservation

The level of the measured tail used for BRIR simulation may be very different from the level in the modeled tail. Thus, simply concatenating the modeled early part with the measured tail may lead to inadequate sound energy. Energy preservation can be done by scaling measured tail up or down depending on the energy level of the modeled tail. Equation 3 shows the calculation of scaling factor *a*, where *t* refers to the concatenation time, *T* refers to the length of the BRIR, p_{mod} is the modeled BRIR and p_{meas} is the measured one.

$$a = \frac{\int_t^T p_{\mathrm{mod}}^2(t)\,dt}{\int_t^T p_{meas}^2(t)\,dt}. \quad (3)$$

During the scaling of measured and modeled BRIRs, left and right channel energy ratio is considered because of the interaural level differences (ILD) of the simulated BRIR.

3.3.2. Cross-fading

Cross-fading is used to prevent abrupt change at the concatenation. For that purpose, a triangular window with length of 512 samples was constructed. A right half of the window is applied to the end of the early part of the modeled BRIR while a left half is applied to the beginning of the measured reverberation tail. Overlap of 256 samples is achieved.

4. Results

Measured, modeled and four simulated BRIRs (20, 40, 60 and 80ms time of concatenation) are included in the analysis. All of them are observed in time domain (decay curves), and in frequency domain (frequency responses). In addition, room acoustic parameters are obtained and compared.

4.1. Time domain

Decay curves of all BRIRs are calculated using the Schroeder backward integration implemented in MatLab software package. Comparing the decay curves of the simulated and the modeled BRIRs, decay curves of the simulated BRIRs are closer to the decay curve of the measured BRIR, Figure 3. This is expected due to the fact that certain part of the measured and simulated BRIRs is the same. With the decreasing of concatenation time, which means that bigger part of measured BRIR is used for the simulation, decay curve of simulated BRIR get closer to the decay curve of the measured one. This leaves the modeled BRIR (blue line at Figure 3) the worst case for the first 100ms of decay curve, while even the adding of measured tail at 80ms makes the decay curve a significantly better approximation of the sound decay in the room.

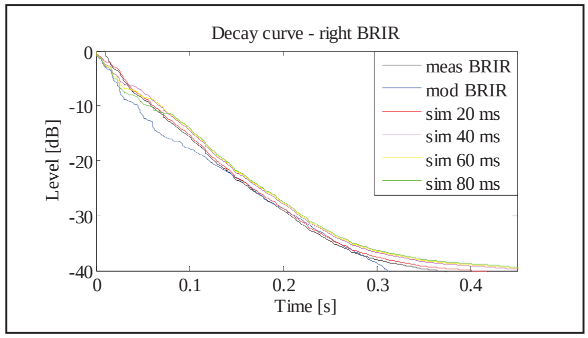

Figure 3. Decay curves of measured (black), modeled (blue) and simulated BRIRs

4.2. Frequency domain

Frequency responses from all six BRIRs are obtained by Fast Fourier Transform. The frequency response of the modeled room represents good approximation of the real room in the range of middle frequencies, Figure 4. Frequencies under about 600 Hz and above 4 kHz are less accurate. Also, frequency response from the right ear is more precise due to the fact that the sound source is located 90^{O} to the right of the receiver and more of the direct sound is received. In the other hand, left ear received more reflected energy. A consequence of sound propagation in enclosed at different frequencies is visible.

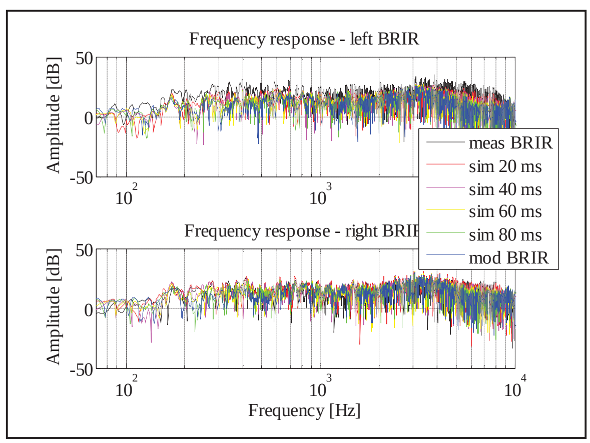

Figure 4. Frequency responses of measured (black), modeled (blue) and simulated BRIRs

Frequency responses obtained by simulated BRIRs, are getting closer to the measured one. This becomes more obvious as the concatenation time is lower. Anyway, even with the concatenation time of 80 ms, closer frequency response ap-

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

proximation is noticed, especially at the lower frequencies. Differences at the higher frequencies are due to the incompatibilities between artificial head used in measurements and the HRTF library used in the modeling process.

Using the measured tail, response from the left ear is much better approximated which can lead to the better approximation of the room acoustic properties. This is noticed even for the short tail (concatenated at 80 ms).

4.3. Room acoustic parameters

Based on the six BRIRs, room acoustic parameters are obtained according to the ISO 3382 normative. Main focus was on the perceptually relevant parameters for the simulation of room acoustics. Literature points at three objective parameters that have most perceptually significant: reverberation time (T), clarity (C) and interaural cross correlation (IACC). These parameters from different BRIRs are compared. Values for the 1kHz octave band are given in Table I.

Table I. Room acoustic parameters for the frequency band with central frequency of 1 kH

BRIR	*T30 left*	*T30 right*	*C50 left*	*C50 right*	*IACCA*
meas	0.423	0.438	3.815	8.033	0.259
mod	0.626	0.617	9.138	11.337	0.331
Sim 20ms	0.430	0.444	4.250	7.527	0.346
Sim 40ms	0.430	0.449	4.020	6.444	0.294
Sim 60ms	0.434	0.464	4.620	7.201	0.281
Sim 80ms	0.454	0.473	6.108	7.837	0.305

Other frequency bands follow the same trend. Comparing simulated and modeled parameters, parameters obtained from the simulated BRIRs are, in general, closer to the measured ones. Still, results for the IACC show certain deviations. Error values are calculated as an absolute difference between "measured" parameters as a reference and parameters obtained from different BRIRs (modeled and simulated). Results are shown at the Figure 5 for the T30, at Figure 6 for the C and at Figure 7 for the IACC.

5. Pilot listening test

Anechoic speech and music signals are convolved with the measured, modeled and simulated BRIRs.

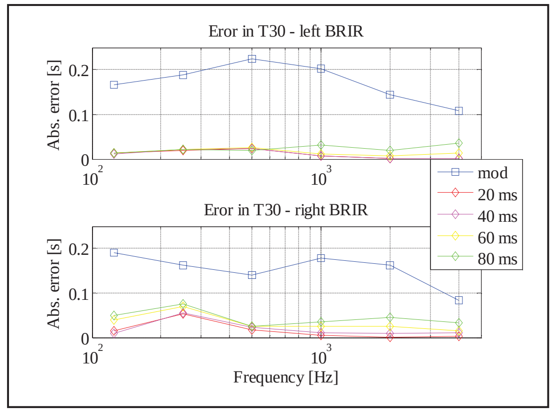

Figure 5. Reverberation time differences

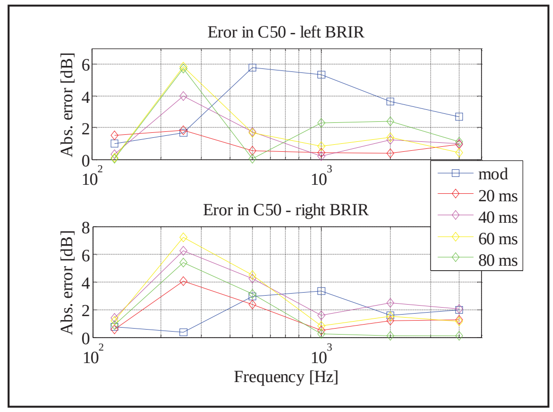

Figure 6. Clarity differences

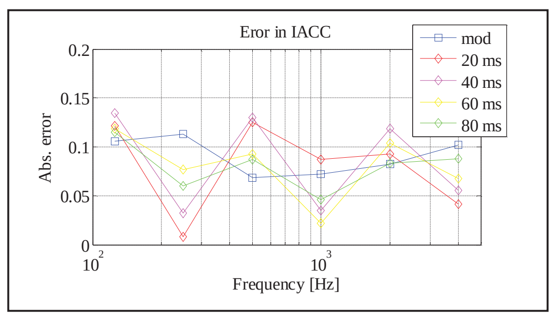

Figure 7. Interaural cross correlation differences

Obtained signals represent static auralization of the standard listening room. Reproduced by headphones, these signals should provide spatial impression of the used anechoic sound. Also, acoustical properties of the room should be accomplished. The listener should have an acoustical feeling of being in the room at the same point where the artificial head used for the measurements were standing.

Auralized signals were compared by the listening test. Subjects were asked whish signal is the most similar to the signal obtained using the measured BRIR.

6. Discussion

Using the measured reverberation tail instead of the modeled one in the binaural synthesis can improve spatial impression and reflect the acoustical properties of existing room more realistic. Even when the measured reverberation tail of 80 ms after the direct sound is used, improvements are noticeable. This is important because of the fact that differences in the BRIRs (obtained from the same room) after 80 ms are not noticeable as the differences in the early part. Thus, this can indicate that finite, relatively small number of measured reverberation tails may improve binaural synthesis without audible lost.
Anyway, further investigation is needed. Listening tests with more subjects have to be done. Also, an investigation about the number of measured reverberation tails that can be used in synthesis has to be performed as a part of the further work.

Acknowledgement

Results presented in this paper are obtained within the FP7 EU collaborative project no. 248620, named “Being in Augmented Multi-Modal Naturally Networked Gatherings” – BEAMING.

References

[1] Y. A. Huang, J Chen, J Benesty: Immersive audio schemes. IEEE Signal Processing Magazine, vol. 28, no. 1, (2011), 20-32.

[2] http://beaming-eu.org/home, official web site of the BEAMING project.

[3] D. R. Begault: 3D sound for virtual reality and multimedia. AP Professional, Cambridge, MA, April 2000.

[4] K. Meesawat, D. Hammershøi: An investigation on the transition from early reflection to a reverberation tail in a BRIR. Proc. 2002 ICAD, 285-289.

[5] J. Blauert: Spatial Hearing, 2nd edition, MIT press, Cambridge, MA, 1997.

[6] K. Meesawat: A study of the Reverberation Tail in Binaural Room Impulse Responses. PhD thesis, Department of Acoustics, Aalborg University, 2004.

[7] H. Kuttruff: Room acoustics, E&FN Spon, 3rd edn, 1991.

[8] J. S. Bradley, G. A. Soulodre: The influence of late arriving energy on spatial impression, JASA, vol. 98, no. 5, (1995), 2263-2271.

[9] http://www.catt.se/, official web site of the CATT-Acoustic software

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767